

QUANTUM INFORMATION: HOW MUCH INFORMATION IN A STATE VECTOR?*

Carlton M. Caves and Christopher A. Fuchs[†]

Center for Advanced Studies, Department of Physics and Astronomy,
University of New Mexico, Albuquerque, New Mexico 87131-1156, USA.

ABSTRACT

Quantum information refers to the distinctive information-processing properties of quantum systems, which arise when information is stored in or retrieved from nonorthogonal quantum states. More information is required to prepare an ensemble of nonorthogonal quantum states than can be recovered from the ensemble by measurements. Nonorthogonal quantum states cannot be distinguished reliably, cannot be copied or cloned, and do not lead to exact predictions for the results of measurements. These properties contrast sharply with those of information stored in the microstates of a classical system.

1. INTRODUCTION

The last fifteen years have seen a steadily increasing exchange of ideas between physicists and information theorists. Physicists have become interested in how modern ideas of information processing affect the physical description of the world around us, and computer scientists and communication theorists have become interested in fundamental questions of how physical law affects information processing. The most fruitful new ideas have arisen from applying information theory to quantum physics, because information in quantum physics is radically different from classical information.

To begin the discussion, we need only the most primitive notion of information content. The fundamental unit of information content, the *bit*, involves two alternatives, conveniently labeled 0 and 1. A bit is not a physical system: it is the abstract unit of information that corresponds

* Supported in part by the U.S. Office of Naval Research (Grant No. N00014-93-1-0116).

[†] Thanks to H. Barnum, R. Jozsa, R. Schack, and B. Schumacher for enlightening discussions. CAF thanks G. Comer in particular for extensive correspondence concerning the point of view set forth here.

to two alternatives; its use implies nothing about how the information is embodied in a physical medium. To transmit ten bits of information from us to you, we assemble a ten-bit message—a string of ten 0s and 1s, say 1101000110—and send it to you. Information thus has to do with selecting one possibility out of a set of alternatives; moreover, information content is a logarithmic measure of the number of alternatives. In the ten-bit example, we select a particular string from the $\mathcal{N} = 2^{10} = 1024$ possible strings, so the information content is $\log_2 \mathcal{N} = 10$ bits. This article, containing nominally about 225 000 bits,¹ corresponding to $2^{225\,000} \simeq 10^{67\,500}$ possible articles, is a more ambitious attempt to transmit information.

One can appreciate the difference between classical and quantum information by comparing and contrasting the physical realizations of a bit in classical and quantum physics. The classical realization of a bit is a classical system that has two possible states—for example, a piece of paper that can have either a 0 or a 1 written on it. To send a bit from us to you, we send you the piece of paper, and you examine it. We provide one bit of information to *specify* which one-bit message to write on the paper or, putting it differently, to *prepare* the appropriate piece of paper. You *acquire* this one bit of information when you examine the paper and determine which message we sent. The paper, while in transit, can be said to *carry* the one bit from us to you. The defining feature of classical information is that when we send an N -bit message, by preparing one of 2^N alternatives, you can acquire all N bits, by distinguishing among the alternatives. This feature is *not* an automatic consequence of physical law. Rather, it is a consequence of using a classical medium to carry the information: in classical physics you are able to distinguish any alternatives we can prepare.

The quantum realization of a bit is a two-state quantum system—for example, a spin- $\frac{1}{2}$ particle. A spin- $\frac{1}{2}$ particle can be used to send one bit of classical information—and no more than one bit—encoded in two orthogonal states, e.g., spin “down” for 0 and spin “up” for 1. Thus it is convenient to denote the state of spin “down” by $|0\rangle$ and the state of spin “up” by $|1\rangle$. The difference between classical and quantum two-state systems is that quantum-mechanical superposition gives a quantum two-state system other possible states, not available to a classical two-state system: any linear combination of $|0\rangle$ and $|1\rangle$ is also a possible state. For a spin- $\frac{1}{2}$ particle these states are in one-to-one correspondence with directions of the particle’s spin.

The crucial distinction between quantum and classical information appears when one attempts to use these other states as alternatives for transmitting information. Suppose we attempt to encode ten bits of information onto a spin- $\frac{1}{2}$ particle, by preparing it so that its spin points in one of 2^{10} possible directions. Then we send the particle to you. Can you read out the ten bits? Of course not. Quantum theory forbids any measurement to distinguish all 1024 possibilities. Indeed, our over-enthusiasm in trying to stuff ten bits into the particle means that the amount

¹ This estimate is obtained by multiplying the number of characters in the T_EX file for this article by 2.14 bits/character; see discussion in Sec. 2.

of information you can recover is considerably less than a bit. Nevertheless, there is a sense in which the spin- $\frac{1}{2}$ particle actually does carry ten bits from us to you; if we want to transmit a description of the particle's state, so that you can prepare another spin- $\frac{1}{2}$ particle with spin in the same direction, we must send you ten classical bits of information. The ten bits of information needed to specify the particle's state, stored in some way in the particle, but not accessible to observation, are an example of quantum information. The fundamental difference between the information-storage and information-retrieval properties of classical and quantum two-state systems has been recognized by dubbing a quantum two-state system a *qubit*.^{2,3}

It is worth repeating the qubit example in a general context. A quantum system can encode classical information in a set of orthogonal states, because orthogonal states can be reliably distinguished by measurements. The number of orthogonal states is limited by the dimension D of the system's Hilbert space, and hence the maximum amount of classical information the system can carry is $\log_2 D$ bits. To emphasize what this means, suppose the system consists of N qubits, where we use $N = 4$ as a running example. The dimension of Hilbert space, $D = 2^N = 16$, is exponentially large in the system size N , and the maximum classical information content, $\log_2 D = N = 4$ bits, is given by the system size.

The superposition principle implies that any linear combination of orthogonal states is a possible state, so the number of possible states of the quantum system is arbitrarily large, limited only by how precisely one specifies the complex quantum-mechanical probability amplitudes for the chosen orthogonal states. Suppose that one gives each amplitude to m -bit accuracy ($m/2$ bits each for the real and imaginary parts), where we use $m = 10$ bits as a running example. Only $D - 1$ amplitudes need be specified, because one amplitude, first made real by a choice of overall phase, is then fixed by normalization. Thus the information needed to specify a state—the quantum information content of the state—is $m(D - 1)$ bits, far larger than the $\log_2 D$ bits of classical information; likewise, the number of quantum states, $2^{m(D-1)}$, is far larger than the Hilbert-space dimension D . For the example of $N = 4$ qubits, the quantum information, $m(D - 1) = m(2^N - 1) = 150$ bits, is exponentially large in system size, and the number of states, $2^{m(D-1)} = 2^{m(2^N-1)} \simeq 10^{45}$, is larger by yet another exponential. *Hilbert space is gratuitously big*—much bigger than the space needed to carry $\log_2 D$ bits. Yet, because no measurement can distinguish all these states, almost none of this huge amount of information is accessible to observation.

² B. Schumacher, "Quantum coding," *Phys. Rev. A* **51**(4), 2738–2747 (1995).

³ *Qubit* is a shorthand for the minimal quantum system, a two-state quantum system, that can carry a bit of information. Logically, if one wishes to give a special name to the minimal physical system that can carry a bit, one should do so for both classical and quantum two-state systems, calling them perhaps c-bits and q-bits. We are reluctant to use the neologism "qubit," because it has no standard English pronunciation, the "qu" being pronounced as in "cue ball," instead of as in "queasy." We prefer "q-bit," but acquiesce in the use of "qubit," which has attained a degree of general acceptance.

How much information is in a state vector? A heck of a lot, but almost none—seemingly a paradox. Fortunately, we stand to gain much from this circumstance: in the words of John Wheeler, “No progress without a paradox!”⁴ The paradoxical character of quantum information is the impetus behind work in the fledgling field of *quantum information theory*, a field seeking to elucidate the nature of quantum information, to quantify it in meaningful ways, and to discover “senses” in which the enormous information content of quantum states can be used.

This article is an introduction to the nature of quantum information. Section 2 begins with the notion of information as having to do with selecting one alternative from an ensemble of possibilities and shows how to quantify the information content of an ensemble in terms of the Gibbs-Shannon information measure. Section 3 sets the stage for the remainder of the article by precisely defining the fundamental alternatives, called “microstates,” for physical systems, in both classical and quantum physics. Classical microstates are fine-grained cells on phase space, and quantum microstates are normalized state vectors, or pure states, in Hilbert space. Section 4 contrasts two measures of the information content of an ensemble of microstates: “preparation information” is the information required to prepare a physical system in a particular microstate drawn from the ensemble, and “missing information” is the information that must be acquired from a measurement to place the system in a microstate. Preparation information and missing information are identical for classical systems, but can be quite different for quantum systems. Sections 5 through 7 focus on three closely related information-theoretic concepts—predictability, distinguishability, and clonability—that strike at the heart of the distinction between ensembles of classical and quantum microstates. Classical microstates lead to precise predictability for all measurements, can be distinguished with certainty, and can be copied or cloned precisely. Non-orthogonal quantum pure states, in contrast, have none of these properties. Section 8 closes the article by noting that these clean information-theoretic distinctions disappear when one compares ensembles of quantum pure states not with ensembles of classical microstates, but with ensembles whose alternatives are overlapping probability distributions for classical microstates. Section 8 thus provides motivation for a longer article, currently in preparation, which explores subtle aspects of quantum information that arise in comparing quantum pure states with classical probability distributions. This article serves as a starting point for the longer paper.

Much of this article is devoted to making a distinction between “maximal” and “complete” information about physical systems. In classical physics maximal information is complete. The distinctive feature of quantum physics is that maximal information is never complete, there being no way to obtain complete information about a quantum system. The distinction between maximal and complete information in quantum physics was brought to the fore in the historic paper of Einstein, Podolsky, and Rosen, a paper that even now, after 60 years, inspires and challenges our thinking. We humbly dedicate this contribution to the memory of Nathan Rosen.

⁴ J. A. Wheeler, “From Mendelée’s atom to the collapsing star,” in *Philosophical Foundations of Science*, edited by R. J. Seeger and R. S. Cohen (Reidel, Dordrecht, 1974), pp. 275–301.

2. INFORMATION CONTENT: A PRIMER

Before proceeding to a comparison of classical and quantum information, we need to sharpen up the primitive notion of information content used in the Introduction, where the information content of \mathcal{N} alternatives was $\log_2 \mathcal{N}$ bits. No probabilities appear here, yet it is clear that they ought to. Suppose, for example, that only two of the alternatives can really occur, those two alternatives being equally likely. There being effectively only two alternatives, the information content of the \mathcal{N} alternatives is only one bit, not $\log_2 \mathcal{N}$ bits.

To get at the notion of information content, we thus must first consider probabilities. Throughout this article we adopt the Bayesian view of probabilities,^{5,6,7,8,9} which holds that a probability is a measure of credible belief based on one's state of knowledge. The Bayesian view is particularly compelling in an information-theoretic context. Probabilities are assigned to a set of alternatives based on what one knows, on one's stock of information about the alternatives. These probabilities are often called *ignorance probabilities*: "One of the alternatives actually occurs, but since I don't know which, all I can do is assign probabilities based on what I do know." Throughout this article we call the existing stock of information, used to make a probability assignment, *prior* information. In defining information content, we seek a measure of *additional* information beyond the prior information that went into assigning the probabilities. What is this additional information? It is the further information, given the prior information, required to *prepare* a particular alternative, or it is the further information, given the prior information, that is *acquired* when one determines a particular alternative.

The task of translating a state of knowledge into a probability assignment is the subject of Bayesian probability theory. Among the chief accomplishments of this theory are a set of rules for assigning prior probabilities in certain cases where the prior information can be given a precise mathematical formulation and the standard rule, called Bayes's rule, for updating a probability assignment as one acquires new information. The general problem of translating a state of knowledge into a probability assignment is, however, far from solved and is the focus of an exciting field of contemporary research.^{8,9,10} That it is not completely solved does not concern us in this article.

Suppose then that there are \mathcal{J} alternatives, labeled by an index j , and that alternative j has probability p_j . We call a collection of alternatives together with their probabilities an

⁵ B. de Finetti, *Theory of Probability*, 2 volumes (Wiley, Chichester, 1974/75).

⁶ L. J. Savage, *The Foundations of Statistics*, 2nd Ed. (Dover, New York, 1972).

⁷ E. T. Jaynes, *Papers on Probability, Statistics and Statistical Physics*, edited by R. D. Rosenkrantz (Kluwer, Dordrecht, 1983).

⁸ J. M. Bernardo and A. F. M. Smith, *Bayesian Theory* (Wiley, New York, 1994).

⁹ E. T. Jaynes, *Probability Theory: The Logic of Science*, to be published.

¹⁰ See also the collection of MaxEnt conference proceedings under the common title *Maximum Entropy and Bayesian Methods* (Kluwer, Dordrecht).

ensemble. To exhibit the role of probabilities in information content, consider what we call a *Gibbs ensemble*, an imaginary construct in which the ensemble of alternatives is repeated N times, N becoming arbitrarily large. The possible configurations of the Gibbs ensemble are sequences of N alternatives, there being \mathcal{J}^N sequences in all. As N becomes large, the probability for the occurrence frequencies of the various alternatives becomes concentrated on those frequencies that match the probabilities in the original ensemble. Thus we only need to consider those sequences for which the frequency of each alternative matches its probability. The total number of such sequences is given by the multinomial coefficient

$$\mathcal{N} = \frac{N!}{\prod_{j=1}^{\mathcal{J}} (Np_j)!}, \quad (1)$$

and each occurs with probability

$$\mathcal{N}^{-1} = \prod_{j=1}^{\mathcal{J}} p_j^{Np_j}, \quad (2)$$

where Stirling's formula relates (1) to (2) for large N . Hence the information content of the Gibbs ensemble—the information required to prepare a particular sequence or the information acquired when one determines a particular sequence—is $\log_2 \mathcal{N} = NH$, where

$$H \equiv - \sum_{j=1}^{\mathcal{J}} p_j \log_2 p_j \quad (3)$$

is called the *Gibbs-Shannon information*.¹¹ Where it is helpful to indicate explicitly the dependence of the Gibbs-Shannon information on a particular probability distribution, we denote it by $H(\tilde{p})$, the symbol \tilde{p} standing for the entire distribution.

The Gibbs-Shannon information $H(\tilde{p})$ can be *interpreted* as the *average* information content per member of the Gibbs ensemble, i.e., as the average information content of the original ensemble. It is the average information required to *specify* a particular alternative within the original ensemble or the average information *acquired* when one determines a particular alternative within the ensemble. For \mathcal{J} alternatives the Gibbs-Shannon information ranges from zero to $\log_2 \mathcal{J}$. When the prior information determines a particular alternative, one assigns unit probability to that alternative and zero probability to all the rest, which leads to $H = 0$; this is the sensible result that when one alternative is known definitely to occur, no information is acquired

¹¹ C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.* **27**, 379–423 (1948) (Part I); **27**, 623–656 (1948) (Part II). Reprinted in book form, with postscript by W. Weaver: C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois, Urbana, IL, 1949).

when it is determined. When the prior information does not discriminate at all among the alternatives, one assigns a uniform probability distribution, which leads to the maximum value of $H = \log_2 \mathcal{J}$; the discussion in the Introduction thus corresponds to assuming minimal prior information and, consequently, a uniform probability distribution. Generally, the Gibbs-Shannon information measures the ignorance that leads to a probability assignment: $H(\tilde{p})$ is the amount of information required to remove the ignorance expressed by the probabilities p_j .

It is interesting to speculate how these considerations affect our estimate that this article conveys 225 000 bits of information. That estimate assumes an average information per letter of 2.14 bits, considerably less than $\log_2(\text{number of different English letters}) \simeq 5$ bits. The reason is that different English letters and combinations of letters occur with markedly different frequencies. The figure of 2.14 bits/letter takes into account correlations between neighboring letters in English text by counting word frequencies for the 8727 most common English words, calculating an information per word, and dividing by an average word length (including spaces) of 5.5 letters to get an information per letter.¹² To estimate the information content of the present text, we multiply the number of characters in the T_EX file by 2.14; in doing so, we are ignoring the difference between letters and T_EX-characters.

This article is correlated from beginning to end; the correlations cannot be ignored in assessing its information content. Even with the nearby correlations of English taken into account, almost all of the $2^{225\,000} \simeq 10^{67\,500}$ alternative articles corresponding to 225 000 bits are gibberish over scales longer than a few words. There is zero probability that we would compose them, and the reader—more importantly, the editors!—would assign zero probability for them to appear in this volume. Of the remaining, much smaller amount of information, some conveys the essential ideas, but most has to do with our style and with attempts to make the essential ideas accessible. In any case, we *must* leave to the reader the delicate task of estimating the essential information conveyed by this article, for that depends critically on the reader’s prior knowledge.

An example of making information accessible, at the cost of redundancy, is a pause to summarize: in the Bayesian view a probability assignment incorporates what one knows about a set of alternatives; the Gibbs-Shannon information quantifies the additional information, beyond what one already knows, to pin down a particular alternative.

¹² C. E. Shannon, “Prediction and entropy of printed English,” *Bell Syst. Tech. J.* **30**, 50–64 (1951).

3. MICROSTATES AND STATES

The remainder of this article formulates differences between classical and quantum information. We draw sharp distinctions between information in classical physics and information in quantum physics by translating into information-theoretic language things physicists already know. We deliberately make the discussion detailed, the risk of tedium being outweighed, we hope, by two benefits. First, many physicists are unfamiliar with information theory and, hence, have trouble appreciating information-theoretic concepts; how better to gain an appreciation of information theory than to see it in action on familiar ground? Second, the sharp distinctions drawn in this article point to subtler questions that arise in distinguishing classical probability distributions from quantum state vectors; these questions, crucial to distinguishing quantum from classical information, are not considered in this article, but are taken up in a subsequent paper.

We are interested in the storage of information in and retrieval of information from physical media and, in particular, in the differences between classical and quantum media. Since information has to do with picking one alternative out of a set of possible alternatives, we must spell out the fundamental alternatives, called *microstates*, in classical and quantum physics.

The arena for classical physics is phase space. At any time a classical system is located at a point in phase space; its dynamics traces out a path through phase space. For specificity, we assume that the system of interest is described on a $2F$ -dimensional phase space, equipped with canonical coördinates $Q_1, \dots, Q_F, P_1, \dots, P_F$; where it is convenient, we abbreviate a phase-space point to $X = (Q_1, \dots, Q_F, P_1, \dots, P_F)$. Furthermore, we assume that the accessible region of phase space has a finite volume

$$\mathcal{V}_F = \mathcal{A}^F, \quad (4)$$

where \mathcal{A} is a typical phase-space area per pair of canonical coördinates.

The fundamental alternatives in classical physics are phase-space points. Yet we cannot specify a typical phase-space point, since the information required to do so is infinite. To keep this information finite, we imagine that there is a finest scale on phase space. This finest scale is characterized by a “resolution volume” $\Delta v_{\text{cl}} = h_0^F \ll \mathcal{V}_F$, where h_0 is a resolution area per pair of canonical coördinates. We grid phase space into fine-grained cells of uniform volume equal to the resolution volume. At this level of fine graining, the fundamental alternatives for a classical system—the classical microstates—are these fine-grained cells. The microstates can be labeled by an index j , and the j th microstate can be specified by the phase-space address $X_j = (Q_1, \dots, Q_F, P_1, \dots, P_F)_j$ of, say, its central point. The number of classical microstates at this level of fine graining is

$$\mathcal{J}_{\text{cl}} = \frac{\mathcal{V}_F}{\Delta v_{\text{cl}}} = \left(\frac{\mathcal{A}}{h_0} \right)^F. \quad (5)$$

Turn now to quantum physics, where the dynamics unfolds within the arena of Hilbert space. More precisely, the relevant space is the space of Hilbert-space rays—normalized state vectors, with vectors that differ by a phase factor considered to be equivalent—a space called *projective Hilbert space*. The dynamics of a quantum system traces out a path on projective Hilbert space. We assume throughout that Hilbert space is finite-dimensional, and we let D denote the number of dimensions.

The fundamental alternatives in quantum physics are normalized state vectors, but the information required to specify a typical state vector is infinite. To keep this information finite, we again imagine that there is a finest resolution, this time on projective Hilbert space. To define a notion of resolution on Hilbert space, we use the natural measure of distance on projective Hilbert space. This natural distance between state vectors $|\psi\rangle$ and $|\psi'\rangle$ is the *Hilbert-space angle*¹³

$$\phi \equiv \cos^{-1}(|\langle\psi|\psi'\rangle|) . \quad (6)$$

Hilbert-space angle translates the overlap between quantum states into a distance function that is derived from a Riemannian metric on projective Hilbert space, called the Fubini-Study metric.^{14,15,16,17} The angle ϕ ranges from zero, when $|\psi'\rangle = e^{i\alpha}|\psi\rangle$, to a maximum value of $\pi/2$, when $|\psi\rangle$ and $|\psi'\rangle$ are orthogonal.

The volume element $d\mathcal{V}_D$ induced by the Fubini-Study metric is put in a form convenient for our purposes by Schack, D'Ariano, and Caves.¹⁸ They choose a fiducial state vector $|\psi_0\rangle$ (analogous to the north pole in three real dimensions) and write an arbitrary normalized state vector as

$$|\psi\rangle = \cos \phi |\psi_0\rangle + \sin \phi |\eta\rangle . \quad (7)$$

Here $\phi \leq \pi/2$ is a “polar” angle, the Hilbert-space angle between $|\psi\rangle$ and $|\psi_0\rangle$, the phase freedom in $|\psi\rangle$ has been removed by choosing $\langle\psi|\psi_0\rangle = \cos \phi$ to be real and nonnegative, and $|\eta\rangle$ is a normalized vector in the $(D-1)$ -dimensional subspace orthogonal to $|\psi_0\rangle$. An integral over projective Hilbert space, i.e., an integral over $|\psi\rangle$, can be accomplished by integrating over the the polar angle ϕ and over the $(2D-3)$ -dimensional sphere of normalized vectors $|\eta\rangle$. The volume element takes the form¹⁸

$$d\mathcal{V}_D = (\sin \phi)^{2D-3} \cos \phi d\phi d\mathcal{S}_{2D-3} , \quad (8)$$

¹³ W. K. Wootters, “Statistical distance and Hilbert space,” *Phys. Rev. D* **23**, 357–362 (1981).

¹⁴ J. Anandan and Y. Aharonov, “Geometry of quantum evolution,” *Phys. Rev. Lett.* **65**(14), 1697–1700 (1990).

¹⁵ J. Anandan, “A geometric approach to quantum mechanics,” *Found. Phys.* **21**(11), 1265–1284 (1991).

¹⁶ G. W. Gibbons, “Typical states and density matrices,” *J. Geom. Phys.* **8**, 147–162 (1992).

¹⁷ S. L. Braunstein and C. M. Caves, “Statistical distance and the geometry of quantum states,” *Phys. Rev. Lett.* **72**, 3439–3443 (1994).

¹⁸ R. Schack, G. M. D'Ariano, and C. M. Caves, “Hypersensitivity to perturbation in the quantum kicked top,” *Phys. Rev. E* **50**(2), 972–987 (1994).

where $d\mathcal{S}_{2D-3}$ is the standard “area” element on a $(2D - 3)$ -dimensional unit sphere. This form of the volume element is most useful for integrands that are symmetric about the fiducial vector and thus depend only on the polar angle ϕ . Integrating over all of projective Hilbert space yields a total volume^{16,18}

$$\mathcal{V}_D = \int d\mathcal{V}_D = \mathcal{S}_{2D-3} \int_0^{\pi/2} d\phi (\sin \phi)^{2D-3} \cos \phi = \frac{\mathcal{S}_{2D-3}}{2(D-1)} = \frac{\pi^{D-1}}{(D-1)!} . \quad (9)$$

Here $\mathcal{S}_{2D-3} = 2\pi^{D-1}/(D-2)!$ is the area of a $(2D - 3)$ -dimensional unit sphere.

To characterize the finest level of resolution on projective Hilbert space, we introduce a quantum “resolution volume” $\Delta v_{\text{qu}} \ll \mathcal{V}_D$. It is convenient to think of these resolution volumes as tiny spheres whose radius, in terms of Hilbert-space angle, is $\phi \ll 1$. The volume of a sphere of radius ϕ is¹⁸

$$\Delta v_{\text{qu}} = \mathcal{S}_{2D-3} \int_0^\phi d\phi' (\sin \phi')^{2D-3} \cos \phi' = (\sin \phi)^{2(D-1)} \mathcal{V}_D \simeq \phi^{2(D-1)} \mathcal{V}_D , \quad (10)$$

where the last form holds for the tiny spheres contemplated here. We assume that the resolution volumes are small enough that sums over resolution volumes can be freely converted to integrals over projective Hilbert space.

At this level of resolution on Hilbert space, the fundamental alternatives—the quantum microstates—are the resolution volumes. The quantum microstates can be labeled by an index j , and the j th microstate can be represented by the state vector $|\psi_j\rangle$ that lies at the center of the j th sphere.¹⁹ A microstate can be specified, for example, by the probability amplitudes $\langle n|\psi_j\rangle$ of the state vector in a specific orthonormal basis $|n\rangle$, $n = 1, \dots, D$, i.e., by the expansion $|\psi_j\rangle = \sum_{n=1}^D |n\rangle \langle n|\psi_j\rangle$. The number of quantum microstates at this level of resolution is²⁰

$$\mathcal{J}_{\text{qu}} = \frac{\mathcal{V}_D}{\Delta v_{\text{qu}}} = \phi^{-2(D-1)} . \quad (11)$$

That there is in practice a finest level of resolution in the description of a physical system follows ineluctably from the inability to store or to process infinite amounts of information. The size of this finest level of resolution, in classical or quantum physics, might be set by indifference to distinctions on yet finer scales or by physical constraints—e.g., by the resolution of laboratory equipment that is available for manipulating the system of interest. It doesn’t really matter how

¹⁹ A sphere is properly represented not by a state vector, but by a uniform distribution of state vectors within the sphere; for small resolution volumes, the difference is unimportant. The analogue in classical physics is that a fine-grained cell is properly represented not by its central point, but by a uniform probability density on the cell.

²⁰ C. M. Caves, “Information, entropy, and chaos,” in *Physical Origins of Time Asymmetry*, edited by J. J. Halliwell, J. Pérez-Mercader, and W. H. Zurek (Cambridge University, Cambridge, England, 1994), pp. 47–89.

the finest scale is chosen; our discussion relies only on choosing a finest scale, not on its actual size.

A microstate is a state at the finest level of description. To say that the system occupies a particular microstate requires that one have *maximal* information about the system. Microstates are thus the states that are determined by maximal information. Though the term “microstate” does convey the notion of a state at the finest level of description, it fails to convey what for us is the more important idea, that of a state specified by maximal information. Nevertheless, lacking a better term, we stick with “microstate” to conform to standard physics terminology.

Microstates: states specified by maximal information

Classical physics	Quantum physics
A microstate is a fine-grained cell	A microstate is a normalized state vector
$X_j = (Q_1, \dots, Q_F, P_1, \dots, P_F)_j \quad (12)$ <p>in phase space.</p>	$ \psi_j\rangle = \sum_{n=1}^D n\rangle \langle n \psi_j\rangle \quad (13)$ <p>in Hilbert space.</p>

What if one does not have maximal information about the system? Then, according to the Bayesian view, one assigns probabilities p_j to the microstates based on what one *does* know. These probabilities quantify ignorance about which microstate the system occupies. The resulting ensemble of microstates and probabilities we call a *state* of the system. The word “state” thus denotes an ensemble in the particular case that the ensemble’s alternatives are microstates of a physical system. We stress that *the system state depends on what one knows about the system*. A microstate is a special case of a state—the special case that is specified by maximal information, i.e., by knowledge of the fundamental alternative. In quantum physics a microstate (or state vector) is often called a *pure state*, whereas an ensemble in which more than one state vector has nonzero probability is called a *mixed state*.

For a classical system the ensemble of microstates and probabilities—the classical state—is equivalent to a phase-space probability density

$$\rho(X) = \sum_{j=1}^{\mathcal{J}_{\text{cl}}} p_j \rho_j(X) , \quad (14)$$

where $\rho_j(X)$ is the normalized uniform density on the j th fine-grained cell. For a quantum system the ensemble of microstates and probabilities—the quantum state—gives rise to a *density operator*

$$\hat{\rho} = \sum_{j=1}^{\mathcal{J}_{\text{qu}}} p_j |\psi_j\rangle \langle \psi_j| = \int d\mathcal{V}_D p(|\psi\rangle) |\psi\rangle \langle \psi| . \quad (15)$$

The last form comes from converting the sum to an integral over projective Hilbert space, with $p(|\psi\rangle)$ being a probability density on projective Hilbert space.

A quantum density operator is sufficient to predict the statistics of all measurements made on the system. Consider, for example, what we call a pure von Neumann measurement, the only kind of measurement considered in this article. A pure von Neumann measurement is a measurement of a nondegenerate observable—i.e., a nondegenerate Hermitian operator—and can be described completely by an orthonormal measurement basis $|n\rangle$, $n = 1, \dots, D$, the eigenbasis of the measured Hermitian operator. (More generally, von Neumann measurements are described in terms of *orthogonal* projection operators, which project onto orthogonal Hilbert-space subspaces; the measurements here are called *pure* because they are described by *one-dimensional* orthogonal projectors $|n\rangle\langle n|$, the projectors onto the orthogonal pure states $|n\rangle$.) Given the ensemble of state vectors $|\psi_j\rangle$ and probabilities p_j , the probability for a pure von Neumann measurement to yield result n is

$$q_n = \sum_{j=1}^{\mathcal{J}_{\text{qu}}} |\langle n|\psi_j\rangle|^2 p_j = \langle n|\hat{\rho}|n\rangle = \text{tr}(\hat{\rho}|n\rangle\langle n|) = \sum_{m=1}^D |\langle n|\phi_m\rangle|^2 \lambda_m. \quad (16)$$

The first form here is a conventional probability formula, since $|\langle n|\psi_j\rangle|^2$ is the conditional probability to obtain result n , given that the system has state vector $|\psi_j\rangle$. The second form in (16) introduces the density operator $\hat{\rho}$ of (15) and shows that it contains the statistics of all pure von Neumann measurements. The third form, with the probability written in terms of a trace, is a form that can be extended to more general kinds of measurements. The last form follows from expanding $\hat{\rho}$ in terms of its own orthonormal eigenbasis $|\phi_m\rangle$,

$$\hat{\rho} = \sum_{m=1}^D \lambda_m |\phi_m\rangle\langle\phi_m|, \quad (17)$$

where λ_m is the eigenvalue of $\hat{\rho}$ associated with the eigenvector $|\phi_m\rangle$. The expansion of a density operator in terms of its own eigenstates and eigenvalues is called its *orthogonal* (or spectral) *decomposition*. The eigenvalues λ_m make up a normalized probability distribution; indeed, if the measurement basis is chosen to be the eigenbasis $|\phi_m\rangle$, λ_m is the probability to obtain result m .

Though the density operator $\hat{\rho}$ is sufficient to predict the statistics of all measurements, it is unlike a classical phase-space density in that it is *not* equivalent to the system state, i.e., to the ensemble of microstates and probabilities. Many different ensembles give rise to the same density operator. Hughston, Jozsa, and Wootters²¹ have outlined a procedure for constructing all ensembles that lead to a given density operator. The lack of equivalence between states and density operators is particularly important when a system can be divided into subsystems.

²¹ L. P. Hughston, R. Jozsa, and W. K. Wootters, “A complete classification of quantum ensembles having a given density matrix,” *Phys. Lett. A* **183**, 14–18 (1993).

Suppose, for example, that a system is made up of two subsystems, and suppose that, having maximal information about the composite system, we assign it a state vector $|\Psi\rangle$. This joint state vector can be expanded as

$$|\Psi\rangle = \sum_m \sqrt{\lambda_m} |\phi_m\rangle |\eta_m\rangle, \quad (18)$$

where the state vectors $|\phi_m\rangle$ are orthogonal state vectors of subsystem 1 and the state vectors $|\eta_m\rangle$ are orthogonal state vectors of subsystem 2. This kind of expansion of a joint pure state is called the *Schmidt decomposition*.²²

The statistics of all measurements on one of the subsystems, say subsystem 1, can be derived from the *marginal density operator* for that subsystem,

$$\hat{\rho}_1 \equiv \text{tr}_2(|\Psi\rangle\langle\Psi|), \quad (19)$$

where tr_2 denotes a *partial trace* over subsystem 2. Any operator \hat{O} on the joint system can be expanded in terms of a product basis $|n\rangle|k\rangle$,

$$\hat{O} = \sum_{n,k,n',k'} O_{nk,n'k'} |n\rangle|k\rangle\langle n'|\langle k'|, \quad (20)$$

where the vectors $|n\rangle$ are an orthonormal basis in the Hilbert space of system 1 and the vectors $|k\rangle$ are an orthonormal basis in the Hilbert space of system 2. A partial trace over subsystem 2 yields an operator on subsystem 1, defined by

$$\text{tr}_2(\hat{O}) \equiv \sum_k \langle k|\hat{O}|k\rangle = \sum_{n,n'} \left(\sum_k O_{nk,n'k} \right) |n\rangle\langle n'|. \quad (21)$$

Carrying out the partial trace to find $\hat{\rho}_1$, we arrive at

$$\hat{\rho}_1 = \sum_{m,m'} \sqrt{\lambda_m \lambda_{m'}} |\phi_m\rangle\langle\phi_{m'}| \text{tr}_2(|\eta_m\rangle\langle\eta_{m'}|) = \sum_m \lambda_m |\phi_m\rangle\langle\phi_m|. \quad (22)$$

The states $|\phi_m\rangle$ and the coefficients λ_m are thus the eigenstates and eigenvalues of $\hat{\rho}_1$. The marginal density operator $\hat{\rho}_1$ is a useful tool for deriving the statistics of measurements on subsystem 1, but *there is no justification for regarding $\hat{\rho}_1$ as associated with any specific ensemble of state vectors and probabilities for subsystem 1*. In particular, one should not regard the state of subsystem 1 as being the ensemble of eigenvectors of $\hat{\rho}_1$ with eigenvalue probabilities. The state in this situation is defined at the level of the joint system; there is no state, in our language, for subsystem 1 alone.

²² A. Peres, *Quantum Theory: Concepts and Methods* (Kluwer, Dordrecht, 1993).

At times throughout the rest of this article we use the example of the *uniform ensemble*, for which the probabilities p_j are all the same. For this case, the classical density $\rho(X) = 1/\mathcal{V}_F$ is uniform on the accessible region of phase space. Similarly, the probability density $p(|\psi\rangle) = 1/\mathcal{V}_D$ is uniform on projective Hilbert space, and by the symmetry of this ensemble, the quantum density operator (15) is a multiple of the unit operator $\hat{1}$ on the D -dimensional Hilbert space:

$$\hat{\rho} = \int \frac{d\mathcal{V}_D}{\mathcal{V}_D} |\psi\rangle\langle\psi| = \frac{\hat{1}}{D}. \quad (23)$$

This ensemble has the virtue of highlighting most dramatically the distinctions between classical and quantum information.

4. PREPARATION INFORMATION AND MISSING INFORMATION

For the remainder of this article, we contrast classical and quantum information by investigating the storage of information in and retrieval of information from classical and quantum systems. The conceit we adopt is the one used in the Introduction: *we* prepare a system in a particular microstate drawn from an ensemble of microstates labeled by j , which have probabilities p_j , and then *we* send the system to *you* for your examination. *We* know which microstate the system occupies; *we* must provide information to prepare the system in that microstate. *You*, not knowing which state we prepared, ascribe to the system the state described by microstate probabilities p_j .

The average amount of information we must provide to *prepare* the system in a particular microstate is the Gibbs-Shannon information $H(\tilde{p})$ corresponding to the probabilities p_j . This is also the amount of information required to *specify* a particular microstate within the ensemble of microstates. We call such information *preparation information*, or *specification information*, and denote it by I . We stress that this preparation information is *not* the information needed to prepare the ensemble, i.e., the system state that has microstate probabilities p_j ; rather, given the system state, it is the average information needed to prepare or to specify a particular microstate within the ensemble. Classically the preparation information can be written as an integral over phase space, and quantum mechanically it can be written as an integral over projective Hilbert space:

$$I = H(\tilde{p}) = - \sum_j p_j \log_2 p_j = \begin{cases} - \int d\mathcal{V}_F \rho(X) \log_2(\rho(X) \Delta v_{\text{cl}}) , & \text{classical,} \\ - \int d\mathcal{V}_D p(|\psi\rangle) \log_2(p(|\psi\rangle) \Delta v_{\text{qu}}) , & \text{quantum.} \end{cases} \quad (24)$$

How might we prepare a particular microstate? One way, which works both classically and quantum mechanically, is to start with the system in a standard microstate and then to apply

a specially designed Hamiltonian that causes the system to evolve into the desired microstate over a specified time interval. For this purpose one can imagine a complicated apparatus that manipulates the system. This preparation apparatus has a dial, whose settings correspond to the system microstates. Setting the dial to the j th microstate adjusts the system Hamiltonian to the designer Hamiltonian that causes the system to evolve into the j th microstate. The setting for the j th microstate is to be used with probability p_j . The information we must provide to pick a particular dial setting is the preparation information I . Parkins *et al.*²³ have proposed an example of this sort of procedure for preparing state vectors of an electromagnetic field mode (a harmonic oscillator).

The designer-Hamiltonian method of state preparation highlights a crucial feature of the preparation information. To prepare the system in a particular microstate, we use another system—the preparation apparatus. The preparation apparatus stores a record—its dial setting—of the prepared microstate. *We*, knowing the dial setting, have maximal information and thus assign the system a microstate; *you*, not knowing the dial setting, but knowing the microstate produced by each dial setting and the probabilities of the settings, ascribe to the system the state described by the possible microstates and their probabilities.

The preparation information should be contrasted with the amount of information that you must acquire from a measurement to obtain maximal information about the system, i.e., to determine a system microstate. This amount of information that you are missing toward a maximal description of the system—*missing information* for short—is, as we see shortly, the *entropy* S of the state, measured in bits.

For a classical system there is no difference between preparation information and missing information. You can make a measurement with the resolution of the fine-grained cells and thereby determine which fine-grained cell the system occupies. The average information you acquire in such a measurement is $H(\tilde{p})$, since microstate j occurs with probability p_j . The missing information is the same as the information we provided to prepare the system—the preparation information—and both are the same as the classical *entropy* of the ensemble [cf. (24)].

The contrast emerges in quantum physics. If the ensemble of microstates includes, with nonzero probability, state vectors that are nonorthogonal, no measurement you make can determine which state vector we prepared. In spite of this, a measurement *can* provide you maximal information about the system, but the state vector you ascribe to the system after your measurement—the one for which the measurement provided maximal information—might not be included in the original ensemble. Here we are thinking about a pure von Neumann measurement,

²³ A. S. Parkins, P. Marte, P. Zoller, and H. J. Kimble, “Synthesis of arbitrary quantum states via adiabatic transfer of Zeeman coherence,” *Phys. Rev. Lett.* **71**(19), 3095–3098 (1993).

described by an orthonormal measurement basis $|n\rangle$, $n = 1, \dots, D$; a pure von Neumann measurement provides maximal information by *leaving* the system in the basis state $|n\rangle$ corresponding to the result n of the measurement.

Given the ensemble of state vectors $|\psi_j\rangle$ and probabilities p_j , the probability q_n for your pure von Neumann measurement to yield result n is given by (16). The average information you acquire from such a measurement is the Gibbs-Shannon information $H(\tilde{q})$ corresponding to the probabilities q_n . The information you acquire clearly depends on the measurement basis $|n\rangle$. How then are we to identify a unique measure of missing information in quantum physics? To do so, we return to our original question: how much information *must* you acquire to obtain maximal information about the system? Thus we seek the von Neumann measurement that yields the *minimum* amount of information, for you must acquire at least this minimal information to obtain a maximal description of the system.

To determine this quantum measure of missing information, we invoke a special property of the quantum conditional probabilities $|\langle n|\phi_m\rangle|^2$ that appear in (16). The conditional probability $|\langle n|\phi_m\rangle|^2$ has a dual character: for a measurement in basis $|n\rangle$, it is the probability to obtain result n , given that the system has state vector $|\phi_m\rangle$, and for a measurement in basis $|\phi_m\rangle$, it is the probability to obtain result m , given that the system has state vector $|n\rangle$. The special property²² we need—dubbed, mysteriously, *double stochasticity*—is a straightforward consequence of this dual character and is simply that these conditional probabilities are normalized on both indices, m and n ,

$$\sum_{n=1}^D |\langle n|\phi_m\rangle|^2 = 1 = \sum_{m=1}^D |\langle n|\phi_m\rangle|^2. \quad (25)$$

Given this, we can write

$$\begin{aligned} H(\tilde{q}) - H(\tilde{\lambda}) &= - \sum_{m,n} |\langle n|\phi_m\rangle|^2 \lambda_m \log_2 \frac{q_n}{\lambda_m} \\ &= \frac{1}{\ln 2} \sum_{m,n} |\langle n|\phi_m\rangle|^2 \left(-\lambda_m \ln \frac{q_n}{\lambda_m} + (q_n - \lambda_m) \right) \geq 0, \end{aligned} \quad (26)$$

where double stochasticity is used to insert the sum over $q_n - \lambda_m$, and where the inequality follows from the property $-\ln x \geq 1 - x$, for which equality holds if and only if $x = 1$. Equality holds in (26) if and only if every term in the second sum vanishes, i.e., $q_n = \lambda_m$ or $\langle n|\phi_m\rangle = 0$ for all n and all m . This necessary and sufficient condition for equality is equivalent to $(\lambda_m - q_n)\langle \phi_m|n\rangle = 0$ for all n and m , which in turn is equivalent to

$$0 = \sum_m (\lambda_m - q_n) |\phi_m\rangle \langle \phi_m|n\rangle = (\hat{\rho} - q_n)|n\rangle \quad \text{for all } n. \quad (27)$$

Thus equality holds in (26) if and only if the measurement basis $|n\rangle$ is an eigenbasis of $\hat{\rho}$.

What we have shown is that^{22,24}

$$H(\tilde{q}) \geq H(\tilde{\lambda}) = - \sum_{m=1}^D \lambda_m \log_2 \lambda_m = -\text{tr}(\hat{\rho} \log_2 \hat{\rho}) \equiv S(\hat{\rho}) , \quad (28)$$

where $S(\hat{\rho})$, the *von Neumann entropy* of the state (or ensemble), is determined by the density operator $\hat{\rho}$. The von Neumann entropy plays a special role: it is the missing information—the minimum amount of information missing toward specification of a microstate—for any ensemble that has density operator $\hat{\rho}$. The measurement that yields this minimum amount of information is a measurement in an eigenbasis of $\hat{\rho}$. The von Neumann entropy ranges from zero, for a pure state, to a maximum of $\log_2 D$, for the density operator $\hat{\rho} = \hat{1}/D$.

In both classical and quantum physics, missing information, or entropy, S is the amount of information missing toward a maximal description of the system. If $S = 0$, there is no missing information; one already knows which microstate the system occupies. When S is greater than zero, there is a measurement that, by acquiring average information S , leaves the system in a microstate; no measurement that provides less information than S can leave one with maximal information about the system.²⁵

We are now ready to appreciate the difference between preparation information and entropy in quantum physics.^{20,26} It is obvious that the preparation information can be much bigger than the entropy. The maximum value of the entropy is determined by the dimension D of Hilbert space, i.e., by the number of *orthogonal* vectors that Hilbert space can accommodate, whereas the maximum value of the preparation information is determined by the number of state vectors, \mathcal{J}_{qu} , in projective Hilbert space. The number of state vectors is much larger than the number of orthogonal vectors, because any superposition of orthogonal state vectors is another state vector, and is limited only by one's resolution on projective Hilbert space; there is no corresponding situation in classical physics, because there is no way to combine two or more fine-grained cells to produce yet another fine-grained cell.

²⁴ A. Wehrl, “General properties of entropy,” *Rev. Mod. Phys.* **50**(2), 221–259 (1978).

²⁵ These ideas are certainly not new with us. Pauli, for instance, had similar thoughts on entropy: “The first application of the calculus of probabilities in physics, which is fundamental for our understanding of the laws of nature, is the general statistical theory of heat, established by Boltzmann and Gibbs. This theory, as is well known, led necessarily to the interpretation of the entropy of a system as a function of its state, which, unlike the energy, depends on our *knowledge* about the system. If this knowledge is the maximal knowledge which is consistent with the laws of nature in general (micro-state), the entropy is always null.” Quotation from W. Pauli, “Probability and physics,” *Dialectica* **8**, 112–124 (1954); translated in W. Pauli, *Writings on Physics and Philosophy*, edited by C. P. Enz and K. von Meyenn (Springer, Berlin, 1994).

²⁶ C. M. Caves, “Information and entropy,” *Phys. Rev. E* **47**(6), 4010–4017 (1993).

The formal statement of the discrepancy between preparation information and entropy is that the former is never smaller than the latter,^{21,24,27,28}

$$I = H(\tilde{p}) = - \sum_{j=1}^{\mathcal{J}_{\text{qu}}} p_j \log_2 p_j \geq -\text{tr}(\hat{\rho} \log_2 \hat{\rho}) = S(\hat{\rho}) . \quad (29)$$

Proofs of (29) can be found in Refs. 21, 24, and 28. Equality holds in (29) if and only if all the state vectors that have nonzero probability are orthogonal—i.e., if and only if the state that gives rise to $\hat{\rho}$ is an ensemble of eigenstates of $\hat{\rho}$ with the probability of each eigenstate given by its eigenvalue.

Preparation information vs. entropy

Classical physics

The amount of information, I , required to prepare a particular microstate within an ensemble of microstates is the same as the entropy S of the ensemble. For the uniform ensemble the amount of preparation information (or entropy) is

$$\begin{aligned} I &= \log_2 \mathcal{J}_{\text{cl}} \\ &= \log_2(\mathcal{V}_F / \Delta v_{\text{cl}}) \\ &= F \log_2(\mathcal{A} / h_0) = S . \end{aligned} \quad (30)$$

We can interpret $\log_2(\mathcal{A} / h_0)$ as the number of bits of preparation information per pair of canonical coördinates.

Quantum physics

The amount of information, I , required to prepare a particular microstate within an ensemble of microstates is larger than the von Neumann entropy S of the ensemble, unless the ensemble consists of orthogonal state vectors. For the uniform ensemble the amount of preparation information is

$$\begin{aligned} I &= \log_2 \mathcal{J}_{\text{qu}} \\ &= \log_2(\mathcal{V}_D / \Delta v_{\text{qu}}) \\ &= (D - 1) \log_2 \phi^{-2} \gg \log_2 D = S(\hat{\rho}) . \end{aligned} \quad (31)$$

We can interpret $\log_2 \phi^{-2}$ as the number of bits of preparation information per probability amplitude (cf. discussion in the Introduction; for the example there of 10 bits per amplitude, $\phi = 1.79^\circ$).

²⁷ L. B. Levitin, “On the quantum measure of information,” in *Proceedings of the Fourth All-Union Conference on Information and Coding Theory*, Sec. II (Tashkent, 1969) (translation available from A. Bezinger and S. L. Braunstein). This paper has been essentially reprinted as a part of L. B. Levitin, “Physical information theory Part II: Quantum systems,” in *Workshop on Physics and Computation: PhysComp '92*, edited by D. Matzke (IEEE Computer Society, Los Alamitos, CA, 1993), pp. 215–219.

²⁸ C. M. Caves and P. D. Drummond, “Quantum limits on bosonic communication rates,” *Rev. Mod. Phys.* **66**(2), 481–537 (1994).

It is instructive at this point to compare directly the number of microstates for a system described in classical physics with the number of microstates for the *same* system described in quantum physics. To do so, imagine fine graining classical phase space on the quantum scale by choosing the resolution area per pair of canonical coördinates, h_0 , to be the Planck constant h . The resulting number of quantum-scale fine-grained cells is $\mathcal{J}_{\text{cl}} = \mathcal{V}_F/h^F$. If such a system is sufficiently classical, i.e., $\mathcal{J}_{\text{cl}} \gg 1$, then when the system is quantized, these quantum-level phase-space cells correspond roughly to *orthogonal* state vectors that span Hilbert space. The number of quantum-level phase-space cells thus gives the dimension of Hilbert space, $\mathcal{J}_{\text{cl}} = \mathcal{V}_F/h^F = D$. The number of quantum microstates, \mathcal{J}_{qu} , is exponentially larger,

$$\mathcal{J}_{\text{cl}} = D \ll 2^{(D-1) \log_2 \phi^{-2}} = \mathcal{J}_{\text{qu}} , \quad (32)$$

as a direct consequence of quantum superposition: superposition of quantum-level phase-space cells produces an exponentially large number of state vectors that have no classical counterpart. Notice that this conclusion is true even if the resolution on projective Hilbert space is so coarse that it corresponds to giving only one bit per amplitude, i.e., $\log_2 \phi^{-2} = 1$.

One can see how quantum statistical physics manages to reduce to classical statistical physics in the classical limit, despite the far larger number of quantum microstates. Statistical physics is founded on entropy, or missing information, not on preparation information. For a sufficiently classical system, the quantum density operator $\hat{\rho} = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ is approximately diagonal in an *orthonormal* basis of state vectors $|\psi_j\rangle$ that can be identified with quantum-level cells on classical phase space. The corresponding classical phase-space density is $\rho(X) = \sum_j p_j \rho_j(X)$, where $\rho_j(X)$ is the normalized uniform density on the j th quantum-level cell. Thus the von Neumann entropy reduces to the classical entropy, provided that the resolution on phase space is fixed at the quantum scale.

The information $H(\tilde{q})$ you acquire from a pure von Neumann measurement provides the information you need to specify the system's state vector *after* the measurement, but for nonorthogonal ensembles the information you acquire is not sufficient to infer the state vector *before* the measurement, i.e., the state vector that we prepared. For nonorthogonal ensembles, part of the information you acquire comes from the intrinsic unpredictability of quantum physics and tells you nothing about which state vector we prepared. Indeed, for nonorthogonal ensembles, this useless part of the information is always large enough that the part that is useful in determining which state vector we prepared, called the *accessible* information, is not just less than the preparation information $H(\tilde{p})$, but is less than the von Neumann entropy $S(\hat{\rho})$.

At this point it is instructive to note that measurements themselves provide another method of state preparation: observe a system, and thereby prepare it in the microstate corresponding to the result of the measurement. In contrast to the use of designer Hamiltonians, however, the state prepared by this method cannot be predicted in advance. In considering preparation by

measurements, assume for simplicity that the state of the system to be observed is the uniform ensemble. Classically, one measures which microstate the system occupies, thereby gathering the $\log_2 \mathcal{J}_{\text{cl}} = S$ bits of missing information (or entropy), which coincides with the preparation information. Things are different in quantum physics. For the uniform ensemble, the density operator is a multiple of the unit operator. Any orthonormal basis is an eigenbasis of this density operator, and thus any pure von Neumann measurement gathers $\log_2 D = S(\hat{\rho})$ bits of missing information. The rest of the preparation information, $\log_2(\mathcal{J}_{\text{qu}}/D)$ bits, comes not from the random measurement outcome, but from the selection of the measurement basis. In contrast to classical physics, only a small part of the preparation information comes from observing the quantum system; most comes from choosing *how* to observe the system.²⁹

So far we have seen that preparation information and missing information are the same in classical physics, but can be quite different in quantum physics. The difference is connected to the fundamental lack of predictability and distinguishability in quantum physics. Our objective is to sharpen up this connection by addressing three closely related questions, concerned with *predictability*, *distinguishability*, and *clonability*.³⁰ These questions are posed in terms of our conceit: *we* prepare the system in a microstate drawn from an ensemble of microstates and probabilities and send the system to *you*; not knowing which microstate we prepared, *you* attribute to the system the state corresponding to the ensemble.

5. PREDICTABILITY

The first question concerns *predictability*: when one has maximal information about a system, do all measurements have predictable results? The question, expressed in our conceit, becomes the following: *we* prepare the system in a microstate from the ensemble of microstates and probabilities and send the system to *you*; can *we* predict uniquely the result of any measurement *you* perform? In both classical and quantum physics, the answer is easy. For a classical system, if one knows the system's microstate, i.e., knows which fine-grained cell the system occupies, then one can predict the results of all measurements made on scales coarser than the chosen fine graining. Measurements yield no new information. In contrast, the essence of quantum physics is that even if one knows the system's microstate, i.e., knows its state vector, unpredictability remains. The outcomes of most measurements are unpredictable and thus yield new information.

²⁹ More elaborate methods for preparing state vectors by measurements have been considered by K. Vogel, V. M. Akulin, and W. P. Schleich, "Quantum state engineering" *Phys. Rev. Lett.* **71**(12), 1816–1819 (1993), and by B. M. Garraway, B. Sherman, H. Moya-Cessa, P. L. Knight, and G. Kurizki, "Generation and detection of nonclassical field states by conditional measurements following two-photon resonant interactions," *Phys. Rev. A* **49**(1), 535–547 (1994).

³⁰ The question of cloning, or copying, state vectors was first considered by W. K. Wootters and W. H. Zurek, "A single quantum cannot be cloned," *Nature* **299**, 802–803 (1982), and independently by D. Dieks, "Communication by EPR devices," *Phys. Lett. A* **92**(6), 271–272 (1982).

A convenient way to quantify the amount of new information was introduced by Wootters.³¹ Given a particular state vector, pick at random a pure von Neumann measurement, and calculate the average information obtained from the measurement, the average being taken over the random choice of measurement. It is equivalent to reverse the roles of the state vector and the measurement³¹: start with a particular pure von Neumann measurement, described by an orthonormal measurement basis $|n\rangle$, pick a random state vector, and then calculate the average information obtained from the measurement, the average being taken over the uniform ensemble of state vectors.

If the state vector is $|\psi\rangle$, the probability to obtain result n is $|\langle n|\psi\rangle|^2$, and the information obtained from the measurement is

$$H = - \sum_{n=1}^D |\langle n|\psi\rangle|^2 \log_2 |\langle n|\psi\rangle|^2. \quad (33)$$

Averaging over the randomly chosen vector $|\psi\rangle$ yields an average information

$$\bar{H} = - \sum_{n=1}^D \int \frac{d\mathcal{V}_D}{\mathcal{V}_D} |\langle n|\psi\rangle|^2 \log_2 |\langle n|\psi\rangle|^2. \quad (34)$$

Every term in the sum is the same, since the integral is independent of the basis vector $|n\rangle$. Replacing $|n\rangle$ with a fiducial vector $|\psi_0\rangle$ and using the volume element of (8), we can write the average information as

$$\begin{aligned} \bar{H} &= -D \int \frac{d\mathcal{V}_D}{\mathcal{V}_D} |\langle \psi_0|\psi\rangle|^2 \log_2 |\langle \psi_0|\psi\rangle|^2 \\ &= -D \frac{\mathcal{S}_{2D-3}}{\mathcal{V}_D} \int_0^{\pi/2} d\phi (\sin \phi)^{2D-3} \cos^3 \phi \log_2(\cos^2 \phi) \\ &= -\frac{D(D-1)}{\ln 2} \int_0^1 dx (1-x)^{D-2} x \ln x. \end{aligned} \quad (35)$$

The final integral can be done by writing $x^s \ln x = (d/ds)x^s$, i.e.,

$$\int_0^1 dx (1-x)^{D-2} x \ln x = \frac{d}{ds} \left(\int_0^1 dx (1-x)^{D-2} x^s \right) \Big|_{s=1} = -\frac{1}{D(D-1)} \sum_{k=2}^D \frac{1}{k}, \quad (36)$$

³¹ W. K. Wootters, “Random quantum states,” *Found. Phys.* **20**(11), 1365–1378 (1990).

thus yielding an average information^{32,33}

$$\bar{H} = \frac{1}{\ln 2} \sum_{k=2}^D \frac{1}{k} . \quad (37)$$

For a two-dimensional Hilbert space, the average information, $\bar{H} = 1/2 \ln 2 = 0.721$ bits, should be compared with the maximum of 1 bit that can be obtained from a pure von Neumann measurement. Similarly, for a three-dimensional Hilbert space, the average information, $\bar{H} = 5/6 \ln 2 = 1.202$ bits, should be compared to the maximum of $\log_2 3 = 1.585$ bits. For large D , the asymptotic value of the average information is

$$\bar{H} \underset{D \gg 1}{\sim} \log_2 D - \frac{1 - \gamma}{\ln 2} = \log_2 D - 0.60995 , \quad (38)$$

where $\gamma = 0.57722$ is Euler's constant; this is just 0.610 bits shy of the maximum of $\log_2 D$ bits that can be obtained from a pure von Neumann measurement. In other words, even when one possesses maximal information about a quantum system, the result of a typical pure von Neumann measurement is nearly completely unpredictable; the measurement yields almost the maximum amount of information that can be obtained from a pure von Neumann measurement.

Predictability?

Classical physics

Yes. If one has the preparation information—i.e., one knows which fine-grained cell the system occupies—then one can predict the results of all measurements on scales coarser than the fine-grained cells. The amount of information acquired from any such measurement is zero.

Quantum physics

No. If one has the preparation information—i.e., one knows the system's state vector—one generally cannot predict the result of a measurement. One acquires further information from a typical measurement. For a measurement chosen at random, the average amount of information acquired is

$$\bar{H} = \frac{1}{\ln 2} \sum_{k=2}^D \frac{1}{k} \underset{D \gg 1}{\sim} \log_2 D - 0.60995 \text{ bits.} \quad (39)$$

The unpredictability of quantum physics lays bare one of its great mysteries: one can gather an arbitrarily large amount of information from a quantum system, by making repeated pure von

³² R. Jozsa, D. Robb, and W. K. Wootters, “Lower bound for accessible information in quantum mechanics,” *Phys. Rev. A* **49**, 668–677 (1994).

³³ K. R. W. Jones, “Entropy of random quantum states,” *J. Phys. A* **23**(23), L1247–L1251 (1990); “Riemann–Liouville fractional integration and reduced distributions on hyperspheres,” *J. Phys. A* **24**, 1237–1244 (1991).

Neumann measurements in incompatible bases. Where does all this information come from? A good example³⁴ is provided by a spin- $\frac{1}{2}$ particle whose spin is measured alternately along the z and x axes. Each measurement yields a bit of information; these bits are plucked out of the system as though one were drawing from an inexhaustible well of information.²⁶

How is this gathering of fresh information from repeated measurements different from the fresh information that is acquired when a classical system is examined on successively finer scales? In classical physics, if one knows which fine-grained cell a system occupies, unpredictability is solely a consequence of making measurements on a scale finer than the original fine graining. If one determines which cell the system occupies at the new, finer scale, predictability is restored at that scale. Not so in quantum physics. Information gathered by repeated measurements has nothing to do with determining the system's state vector on finer and finer scales on projective Hilbert space. The new information does not enhance predictability at all. With each new measurement some quantities become more predictable, while others become less predictable, as in the example of the spin- $\frac{1}{2}$ particle.

What does this tell us about the status of probabilities in quantum physics? Consider, for example, the density operator $\hat{\rho}$ of (15) and (17). If one makes a measurement in the eigenbasis of $\hat{\rho}$, the probability to obtain result m is given by the eigenvalue

$$\lambda_m = \langle \phi_m | \hat{\rho} | \phi_m \rangle = \sum_{j=1}^{\mathcal{J}_{\text{qu}}} |\langle \phi_m | \psi_j \rangle|^2 p_j . \quad (40)$$

There appear to be two quite different kinds of probabilities in this expression: the prior probabilities p_j express ignorance about the system's microstate; the conditional probabilities $|\langle \phi_m | \psi_j \rangle|^2$, which give the probability to obtain result m given that the system has state vector $|\psi_j\rangle$, express the intrinsic unpredictability of quantum physics.

One's first inclination is to view the conditional probabilities not as ignorance probabilities, but as something else, say, "quantum probabilities."^{35,36,37,38} These "quantum probabilities" are determined by the rules of quantum physics; i.e., they are squares of Hilbert-space inner products. Like the probabilities $|\langle n | \psi \rangle|^2$ in (33), they can be thought of as conditional probabilities for

³⁴ E. P. Wigner, "On hidden variables and quantum mechanical probabilities," *Am. J. Phys.* **38**(8), 1005–1009 (1970). In a footnote, Wigner attributes discussion of this example to von Neumann, who based his private belief in the inadequacy of hidden-variable theories upon it.

³⁵ J. von Neumann, "Quantum logics (strict- and probability-logics)," in *John von Neumann: Collected Works*, Vol. IV, edited by A. H. Taub (Macmillan, New York, 1962), pp. 195–197.

³⁶ P. Benioff, "Possible strengthening of the interpretative rules of quantum mechanics," *Phys. Rev. D* **7**(12) 3603–3609 (1973).

³⁷ M. Strauss "Two concepts of probability in physics," in *Logic, Methodology and Philosophy of Science IV*, edited by P. Suppes, L. Henkin, A. Jojo, and Gr. C. Moisil (North-Holland, Amsterdam, 1973), pp. 603–615.

³⁸ K. R. Popper, *Quantum Theory and the Schism in Physics* (Hutchinson, London, 1982).

measurement results, conditioned on a particular state vector, i.e., on a maximal description of the system. How can they be expressions of ignorance, when they follow from maximal information and the very laws of physics?

Initial enthusiasm for having two different species of probabilities is dampened by the realization that they get hopelessly entangled, leaving no way to maintain a clean distinction. In (40) the probabilities λ_m of the measurement outcomes are a combination of the purported species of probabilities. The form of the combination depends on the ensemble, even though the λ_m themselves, the eigenvalues of $\hat{\rho}$, are blind to the ensemble that defines the density operator. For example, if the ensemble consists of the eigenstates of $\hat{\rho}$, with probabilities λ_m , then these probabilities are purely ignorance probabilities, whereas if $\hat{\rho}$ is constructed from a nonorthogonal ensemble, the same probabilities λ_m contain contributions from “quantum probabilities.”

We suggest that the best approach is to adhere to the Bayesian view, which holds that *all* probabilities are expressions of ignorance about a set of alternatives. An argument³⁹ against this approach—against viewing “quantum probabilities” as ignorance probabilities—runs as follows. If probabilities express ignorance, then by removing the ignorance, one should be left with *complete* information, which permits one to predict everything with certainty. In classical physics this is just what happens, but in quantum physics it is not. In classical physics one acquires complete information and complete predictability by determining which microstate the system occupies. In quantum physics there is no procedure for removing all ignorance. One is not allowed to acquire *complete* information; the best one can do is to acquire *maximal* information, which leaves one uncertain about the results of most observations. Thus, the argument runs, since not all ignorance can be removed, “quantum probabilities” cannot be ignorance probabilities.

We reject this argument, because requiring that all ignorance be removable is just a prejudice, not valid in quantum physics. “Quantum probabilities” are ignorance probabilities; they express ignorance about the outcomes of potential measurements. *What is different in quantum physics is not the status of probabilities, but rather the nature of the alternatives.* In classical physics, probabilities are concerned with *actualities*: “One of the alternatives actually occurs, but since I don’t know which, I assign probabilities based on what I do know.” The probabilities that describe intrinsic quantum unpredictability—the “quantum probabilities”—express ignorance about *potentialities*⁴⁰ that are actualized by measurement: “I know one of these alternatives *would* occur if I enquired about that set of alternatives, but since I don’t know which, I assign probabilities based on what I do know.” What is it that one knows? The system’s state

³⁹ For a related argument, see R. N. Giere, “Objective single-case probabilities and the foundations of statistics,” in *Logic, Methodology and Philosophy of Science IV*, edited by P. Suppes, L. Henkin, A. Jojo, and Gr. C. Moisil (North-Holland, Amsterdam, 1973), pp. 467–483.

⁴⁰ W. Heisenberg, “The development of the interpretation of the quantum theory,” in *Niels Bohr and the Development of Physics*, edited by W. Pauli (McGraw-Hill, New York, 1955), pp. 12–29; W. Heisenberg, *Physics and Philosophy: The Revolution in Modern Science* (Harper, New York, 1958).

vector. Given that knowledge, quantum physics provides the rule for assigning probabilities to the results of *all* possible questions that can be addressed to the system. The sometimes perceived weakness of Bayesianism—that there is no general theory for translating a state of knowledge into a probability assignment—does not apply to the case of maximal information in quantum physics. Indeed, viewed in this light, the quantum rule for assigning probabilities is the most powerful rule yet of Bayesian probability theory.

Formally, one says that in classical physics, maximal information is complete, but in quantum physics, it is not.⁴¹ What should we demand of a physical theory in which maximal information is not complete? Maximal information is a state of knowledge; the Bayesian view says that one must assign probabilities based on the maximal information. Classical physics is an example of the special case in which all the resulting probabilities predict unique measurement results; i.e., maximal information is complete. In a theory where maximal information is not complete, the probabilities one assigns on the basis of maximal information are probabilities for answers to questions one might address to the system, but whose outcomes are not necessarily predictable (some outcomes must be unpredictable, else the maximal information becomes complete). This implies that the possible outcomes cannot correspond to actualities, existing objectively prior to asking the question; otherwise, how could one be said to have maximal information?⁴² Furthermore, the theory must provide a rule for assigning probabilities to *all* such questions; otherwise, how could the theory itself be complete? Quantum physics is consistent with these demands. A more ambitious program would investigate whether the quantum rule is the *unique* rule for assigning probabilities in situations where maximal information is not complete. You won't be surprised to learn that we don't know how to make progress on this ambitious program.

With this perspective, let us return to the eigenvalue probabilities λ_m in (40). These probabilities express ignorance about the result of a measurement in the eigenbasis of the density operator $\hat{\rho}$. Different states of knowledge, i.e., different ensembles, can lead to the same density operator and thus to the same eigenvalue probabilities. It is not the status of the probabilities

⁴¹ It is worth noting for historical interest that neither Einstein nor Pauli would have been a stranger to this language, though they were in opposing camps on the foundations of quantum physics; their difference lay in their beliefs about the finality of this situation. In a letter from Pauli to Born, dated 1954 April 15, Pauli says, “What is more, on the occasion of my farewell visit to him [Einstein] he told me what we quantum mechanicians would have to say to make our logic unassailable (but which does *not* coincide with what he himself believes): ‘Although the description of physical systems by quantum mechanics is incomplete, there would be no point in completing it, as the complete description would not agree with the laws of nature.’” From M. Born, *The Born-Einstein Letters* (Walker, New York, 1971), p. 226. Alternatively, in a letter from Pauli to M. Fierz, dated 1954 August 10, Pauli says, “The famous ‘incompleteness’ of quantum mechanics (Einstein) is somehow-somewhere really there, but of course cannot be remedied by going back to the physics of classical fields (this is just a ‘neurotic misunderstanding’ on Einstein’s part.). . .” From K. V. Laurikainen, *Beyond the Atom: The Philosophical Thought of Wolfgang Pauli* (Springer, Berlin, 1988), p. 145 (translation by R. Schack).

⁴² Here we argue that in a theory where maximal information is not complete, the quantities the theory deals with cannot all be actualities, i.e., objective properties. Bell inequalities do something more in the case of quantum physics: they *demonstrate* that quantum physics has no extension to a theory in which maximal information is

λ_m that changes when going from one such ensemble to another; what changes is the nature of the alternatives to which the probabilities apply. For example, if the ensemble consists of the eigenstates of $\hat{\rho}$, with probabilities λ_m , then the alternatives have the properties of actualities, whereas if $\hat{\rho}$ is constructed from a nonorthogonal ensemble, the same alternatives can only be potentialities.

The information gathered from repeated measurements on quantum systems is indeed drawn from an inexhaustible well, but it is a well of potentialities, not actualities. Asked where all this information resides, we reply, with apologies to Gertrude Stein⁴³: “There is no where there.”

6. DISTINGUISHABILITY

Putting aside this philosophical discussion, we return to our series of questions, the second of which concerns *distinguishability*: can microstates be distinguished reliably by measurements? In terms of our conceit, the question becomes the following: *we* prepare the system in a microstate from the ensemble of microstates and probabilities and send the system to *you*; can *you*, not knowing which microstate we prepared, determine the microstate from the result of a measurement?⁴⁴ Again the answer is easy. In classical physics, yes, because a measurement can determine which fine-grained cell we prepared. In quantum physics, no, for nonorthogonal ensembles, because no measurement can distinguish nonorthogonal state vectors reliably.

The problem of trying to determine which microstate we sent—an *inference* problem—is easy to state, but when the inference is not completely reliable, it is difficult to formulate a quantitative measure of just how reliable the inference is.⁴⁵ For this reason it is convenient to replace the inference problem with a related question taken from communication theory: *we* provide information to prepare the system in a microstate drawn from the ensemble of microstates and probabilities; can the preparation information be transmitted from us to *you*? For a given measurement, the amount of information you acquire about which microstate we sent is called

complete, i.e., a theory in which the statistical predictions of quantum physics are an expression of ignorance about underlying actualities, usually called hidden variables. More precisely, the Bell inequalities show that the statistical predictions of quantum physics cannot be obtained from any *local* theory in which maximal information is complete.

⁴³ Stein damned Oakland, California, her childhood home, with the comment, “There is no there there.”

⁴⁴ It is useful to spell out very clearly what is happening here. *We* prepare a system in such a way that we possess maximal information about it; that maximal information specifies a particular microstate from the given ensemble. *We* send the system to you. The question is the following: can *you*, by performing measurements on the system, acquire the maximal information that *we* used to prepare the system? This formulation underscores the fact that your objective is to find out what we did—i.e., which of the possible procedures we used to prepare the system. The system is the medium from which you try to extract the information you desire.

⁴⁵ C. A. Fuchs, *Distinguishability and Accessible Information in Quantum Theory*, Ph.D. thesis, University of New Mexico (1996).

the *mutual information*. The mutual information is the amount of information transmitted from us to you; it cannot exceed the preparation information.

We are interested in your *optimal* measurement, the one that provides the most information about which microstate we sent. The corresponding maximum value of the mutual information, called the *accessible information*⁴⁶ and denoted here by J , provides a quantitative measure of the distinguishability of the microstates in the ensemble. For an ensemble of classical microstates, you can read out all the preparation information by observing which fine-grained cell the system occupies. Similarly, for an ensemble of *orthogonal* state vectors, you can read out all of the preparation information by making a pure von Neumann measurement in a basis that includes the orthogonal states. For an ensemble of *nonorthogonal* state vectors, however, not all of the preparation information can be transmitted from us to you—the accessible information is necessarily less than the preparation information—because no measurement you make can distinguish all the states in the ensemble.

The uniform ensemble illustrates these points without requiring the entire formalism of quantum communication theory. We select a state vector from the uniform ensemble and send it to you. You make a pure von Neumann measurement described by an orthonormal basis $|n\rangle$. It doesn't matter which basis you choose, because any two bases are related by a unitary transformation, which leaves the uniform ensemble unchanged. How much of the preparation information do you obtain? All the measurement outcomes are equally likely, because the density operator is $\hat{\rho} = \hat{1}/D$, so you acquire $\log_2 D$ bits from the measurement. Yet almost all of this information is due to the intrinsic unpredictability of quantum mechanics—i.e., it is information about quantum potentialities that are actualized by the measurement—and thus it is not information about which state vector we sent. Indeed, if you knew which state vector we transmitted, you would still obtain on average the unpredictability information \bar{H} of (37). This portion of the $\log_2 D$ bits you obtain is evidently not information about which state vector we sent, so it must be subtracted from the $\log_2 D$ bits acquired from the measurement, to yield an accessible information

$$J = \log_2 D - \bar{H} = \log_2 D - \frac{1}{\ln 2} \sum_{k=2}^D \frac{1}{k}. \quad (41)$$

For a two-dimensional Hilbert space the accessible information is $J = 1 - 1/2 \ln 2 = 0.279$ bits, for a three-dimensional Hilbert space it is $J = \log_2 3 - 5/6 \ln 2 = 0.383$ bits, and for large D the asymptotic value is

$$J \underset{D \gg 1}{\sim} \frac{1 - \gamma}{\ln 2} = 0.60995 \text{ bits}. \quad (42)$$

⁴⁶ B. Schumacher, "Information from quantum measurements," in *Complexity, Entropy, and the Physics of Information*, edited by W. H. Zurek (Addison-Wesley, Redwood City, CA, 1990), pp. 29–37.

The upshot is that you get almost none of the preparation information. Indeed, the accessible information is smaller than the von Neumann entropy $S(\hat{\rho}) = \log_2 D$ —much smaller for large dimensions—and the von Neumann entropy, in turn, is much smaller than the preparation information, $I = (D - 1) \log_2 \phi^{-2}$, for the uniform ensemble.

Distinguishability?

Classical physics

Yes. Measurements can distinguish different microstates unambiguously. Hence, the preparation information can be transmitted reliably. In the case of the uniform ensemble, the accessible information is

$$J = \log_2 \mathcal{J}_{\text{cl}} = S = I; \quad (43)$$

the accessible information is equal to the classical entropy S and to the preparation information I .

Quantum physics

No. No measurement can distinguish nonorthogonal microstates unambiguously. Hence, not all the preparation information for a nonorthogonal ensemble can be transmitted reliably. In the case of the uniform ensemble, the accessible information is

$$\begin{aligned} J &= \log_2 D - \bar{H} \\ &= \log_2 D - \frac{1}{\ln 2} \sum_{k=2}^D \frac{1}{k} \underset{D \gg 1}{\sim} 0.60995 \text{ bits}; \end{aligned} \quad (44)$$

the information transmitted is smaller than the von Neumann entropy $S(\hat{\rho}) = \log_2 D$ (for $D \gg 1$, much smaller), which in turn is much smaller than the preparation information I .

Jones^{47,48} has formulated and investigated the problem of using measurements to determine which state vector is drawn from the uniform ensemble and has generalized to the case where one is allowed many copies of the same state vector. He replaces the inference problem with the corresponding communication problem, just as we do here, and uses mutual information to characterize the inference power of the measurements. As part of his investigation, he has developed powerful mathematical tools^{33,48} for doing Hilbert-space integrals like (35).

In the example of the uniform ensemble, unpredictability translates directly into lack of distinguishability; i.e., the average information \bar{H} , which quantifies unpredictability, is subtracted from $\log_2 D$ to give the accessible information. Intrinsic quantum unpredictability can be thought of as drawing from the well of information about potentialities. Any attempt to distinguish nonorthogonal states must draw from this well, the resulting information acting as noise that

⁴⁷ K. R. W. Jones, “Principles of quantum inference,” *Ann. Phys. (N.Y.)* **207**(1), 140–170 (1991).

⁴⁸ K. R. W. Jones, “Fundamental limits upon the measurement of state vectors,” *Phys. Rev. A* **50**(5), 3682–3699 (1994).

defeats the attempt. Though one can extract as much information as one wants from a quantum system, by going repeatedly to the well of information about potentialities, one cannot acquire the information needed to distinguish nonorthogonal state vectors.

7. CLONABILITY

The third question on our list concerns *clonability*: can microstates be copied reliably? In terms of our conceit, this question becomes the following: *we* prepare the system in a microstate from the ensemble of microstates and probabilities and send the system to *you*; can *you*, not knowing which microstate we prepared, devise a procedure that, without changing the microstate of the system we sent, prepares a second system in the same microstate.

The answer is easy this time partly because we can argue that clonability is equivalent to distinguishability. Distinguishability implies clonability: if microstates are distinguishable, you can determine which microstate we sent and then, employing an appropriate designer Hamiltonian, prepare a second system in the same microstate. Conversely, clonability implies distinguishability (provided you can determine a microstate if you have an arbitrarily large number of copies of it): if you can prepare one copy of a microstate, you can prepare an arbitrarily large number of copies and thereby determine the state. The assumption here—that you can determine a microstate if you have an arbitrarily large number of copies—is certainly true in classical physics—the copies, though unnecessary, don’t hurt—but it is also true in quantum physics. Given an arbitrarily large number of copies of a state vector, you can identify the state vector by determining the statistics of pure von Neumann measurements in a sufficient number of incompatible bases.^{49,50,51}

For an ensemble of classical microstates, you can observe which fine-grained cell the system occupies and then prepare a second system in the same fine-grained cell. Similarly, for an ensemble of *orthogonal* state vectors, you can determine which state vector we prepared by making a pure von Neumann measurement in a basis that includes the orthogonal states and then prepare a second system in the same state. For an ensemble of *nonorthogonal* state vectors, however, you cannot devise a procedure that copies all the state vectors in the ensemble.

The standard demonstration³⁰ that state vectors generally cannot be cloned runs as follows. Suppose one wishes to make N copies of a state vector $|\psi\rangle$. One starts with the original system having state vector $|\psi\rangle$ and with the N systems that are to receive the copies having some standard

⁴⁹ J. L. Park and W. Band, “A general theory of empirical state determination in quantum physics: Part I,” *Found. Phys.* **1**(3), 211–226 (1971).

⁵⁰ I. D. Ivanović, “Geometrical description of quantal state determination,” *J. Phys. A* **14**(1), 3241–3245 (1981).

⁵¹ W. K. Wootters, “Quantum mechanics without probability amplitudes,” *Found. Phys.* **16**(4), 391–405 (1986).

state vector. A copying transformation takes this initial state vector to a final product state vector in which all $N + 1$ systems have state vector $|\psi\rangle$:

$$|\psi\rangle \left| \begin{array}{c} \text{standard} \\ \text{state vector of} \\ N \text{ systems} \end{array} \right\rangle \xrightarrow[\text{transformation}]{\text{copying}} |\psi\rangle \underbrace{|\psi\rangle \dots |\psi\rangle}_{N \text{ times}}. \quad (45)$$

There is nothing wrong with this transformation for a single state vector $|\psi\rangle$, but problems arise when one tries to copy all the state vectors $|\psi_j\rangle$ in an ensemble. Since the transformation must be unitary, it must preserve the inner product between any two initial state vectors to which it is to be applied. Thus unitarity requires that

$$\langle \psi_j | \psi_k \rangle = \langle \psi_j | \psi_k \rangle \left\langle \begin{array}{c} \text{standard} \\ \text{state vector of} \\ N \text{ systems} \end{array} \middle| \begin{array}{c} \text{standard} \\ \text{state vector of} \\ N \text{ systems} \end{array} \right\rangle = \langle \psi_j | \psi_k \rangle^{N+1}, \quad (46)$$

which is equivalent to $\langle \psi_j | \psi_k \rangle = 0$ or 1 . The unitarity requirement can be met if and only if $|\psi_j\rangle$ and $|\psi_k\rangle$ are identical or orthogonal. An ensemble of orthogonal state vectors can be cloned, but an ensemble of (distinct) nonorthogonal state vectors cannot.

This demonstration that nonorthogonal state vectors cannot be cloned, when combined with the preceding argument that distinguishability implies clonability, seems to show definitively that nonorthogonal state vectors cannot be distinguished. The alert reader, however, will wonder what this demonstration, which relies on unitarity, has to do with the preceding argument, which was couched in terms of measurements; perhaps nonorthogonal states can be cloned if one allows measurements to be part of the process. It's easy to demonstrate that this cannot be the case.³⁰ An apparatus designed to make measurements and to prepare copies can be included in the overall copying transformation, which becomes a grand unitary transformation for the $N + 1$ systems and the apparatus. If the apparatus starts in some standard pure state, the copying transformation must take the form

$$|\psi\rangle \left| \begin{array}{c} \text{standard} \\ \text{state vector of} \\ N \text{ systems} \end{array} \right\rangle \left| \begin{array}{c} \text{standard} \\ \text{state vector of} \\ \text{apparatus} \end{array} \right\rangle \xrightarrow[\text{transformation}]{\text{copying}} |\psi\rangle \underbrace{|\psi\rangle \dots |\psi\rangle}_{N \text{ times}} \left| \begin{array}{c} \text{final} \\ \text{state vector of} \\ \text{apparatus} \end{array} \right\rangle_{|\psi\rangle}, \quad (47)$$

where the final apparatus state vector can depend on $|\psi\rangle$. The crucial feature of the final state—that it is a product state vector of the $N + 1$ systems and the apparatus—is necessary because one desires a unique state vector for the $N + 1$ systems. Carrying through the requirements of unitarity, one finds that including the apparatus does not change the previous conclusion.

Clonability?

Classical physics	Quantum physics
Yes. Microstates can be copied.	No. Microstates in an ensemble of non-orthogonal state vectors cannot be copied.

Why consider clonability separately from distinguishability when the two are equivalent? The easy answer is that it's always good to have equivalent formulations of a problem. In this case one can say more: the question of clonability is somehow easier to formulate than the question of distinguishability. As a result, consideration of clonability allows one to see directly the principle that prevents cloning—and, hence, prevents distinguishing—nonorthogonal state vectors. That principle is the unitarity of quantum physics. No matter how involved the demonstrations of the inability to distinguish nonorthogonal quantum states become, the underlying principle is unitarity.

8. CONCLUSION

This article has been devoted to comparing and contrasting the information storage and retrieval properties of classical and quantum systems. Information is stored by preparing a system in a particular microstate drawn from an ensemble of possible microstates. Information is retrieved by observing the system and trying to determine which microstate was prepared. The properties of classical information—i.e., information encoded in the microstates of a classical system—are a consequence of the distinguishability of classical microstates. Information put into a classical system can be recovered by observing the system. Information stored in orthogonal microstates of a quantum system acts just like classical information, because orthogonal state vectors can be distinguished by measurements. Quantum information has something to do with the information needed to prepare or to specify a particular state vector from an ensemble of nonorthogonal state vectors. Though quantum measurements can generate as much new information as desired, the information used to prepare one state vector from an ensemble of nonorthogonal state vectors is not accessible to observation, because nonorthogonal state vectors cannot be distinguished.

Should we stop here, satisfied with this clean distinction between classical and quantum microstates? No, because the distinction disappears when one compares, instead, classical probability distributions and quantum state vectors. An ensemble whose members are themselves *overlapping* probability distributions has none of the properties of classical information discussed in this article. Indeed, such an ensemble displays the contrasting properties of an ensemble of nonorthogonal state vectors: a probability distribution does not provide predictability, overlapping distributions cannot be distinguished, and overlapping distributions cannot be cloned. Yet an ensemble of probability distributions is a purely classical concept and can have nothing to do

with quantum information. It must be possible—indeed, it is essential if we wish to understand the differences between classical and quantum information—to find clean distinctions between ensembles of probability distributions and ensembles of state vectors. This task, for which this article serves as a prelude, we must defer to a more extensive future paper.